深度学习在时间序列预测任务中的应用综述

黄佳俊　北京化工大学

摘 要

随着传感器和网络技术的飞速发展，大量历史时间序列数据出现，高效准确地进行时间序列预测越来越重要。近年来，将深度学习的思想和技术运用到时间序列预测任务中的方法发展迅速，并取得了许多成果。本文分析了时间序列预测方法的国内外研究现状，论述了时间序列预测所涉及的相关理论，归纳总结了该任务所运用的传统方法、基于机器学习的方法和基于深度学习的方法，重点对比分析了基于深度学习的各方法的优缺点，并由此对基于深度学习的时间序列预测方法进行了展望。

关键词：时间序列预测，深度学习，机器学习

目录

第一章 绪论

1.1 背景与意义

时间序列数据在金融、气象、农业、工业及医疗等领域存在广泛的应用[1-3]。近年来，随着传感器和网络技术的高速发展，使得生成并积累大量的时间序列数据成为了可能。对于时间序列数据领域来说，通常涉及到分类、异常检测和预测等多个研究问题，其中时间序列预测任务是其研究的重点。与普通的回归分析预测模型不同，时间序列模型更依赖于数据在时间维度上的先后顺序。时间序列预测包括连续型预测（数值预测或范围估计）与离散型预测（事件预测）等，其核心就是从过去时序数据中挖掘出规律，并利用其对将来的发展趋势做出估计。由于时间序列预测研究能够分析其蕴含的规律来预测未来发展趋势，为各行业的决策过程提供指导，因此具有很高的学术意义和应用价值。

对于时间序列预测问题来说，随着时序数据积累量增加和数据维度增长，其研究方法也不断得到改进。从最初的数学统计方法进展到机器学习的方法，再发展到深度学习的方法。由于自然语言处理本质上也就有时序性要求，随着深度学习在自然语言处理任务上取得的巨大成果，因此被广泛使用的深度学习技术也被借鉴到时间序列研究领域。基于深度学习的时间序列预测方法得到了快速的发展，因此有必要对其发展的历史和现状进行分析和讨论。

本文首先对于时间序列预测问题进行了讨论，然后对于其研究方法进行了梳理，针对基于深度学习的研究方法按照其不同的实现网络类型分别进行了分析和讨论，在文章最后对于时间序列预测技术的发展进行了进一步展望。

1.2 时间序列

时间序列一般是指对某种事物发展变化过程进行观测并按照一定频率采集得出的一组随机变量。

由于时间序列数据各变量间具有固有的潜在联系，常表现出一种或多种特性。为了更全面地了解时序预测，本章将详细介绍以下常见特性：

1、海量性随着物联网传感器设备的升级、测量频率的提高以及测量维度的增加，

时间序列数据呈爆炸性增长，高维度的时间序列数据占据主流。在数据集层面进行有效的预处理工作是高质量完成时间序列预测任务的关键。

2、趋势性：当前时刻的数据往往与前一段时刻的数据密切相关，这一特点表明时间序列数据受其他因素影响通常有一定的变化规律，可能在长时间内展现出平稳上升、平稳下降或保持水平的趋势。

3、周期性：时间序列中的数据受外界因素影响，在长时间内呈现出起伏的交替变化。例如，涨潮退潮，一周内潮水高度不符合趋势性变化，不是朝着某一方向的近似直线的平稳运动。

4、波动性：随着时间推移和外部多因素的影响，时间序列的方差和均值可能会发生系统性变化，从而在一定程度上影响时间序列预测的准确度。

5、平稳性：时间序列数据在个别时间点上可能表现为随机变动，但在不同时间上呈现统计规律，同时在方差和均值上保持相对稳定。

6、对称性：如果某段时间周期内，原始的时间序列和其反转时间序列的距离控制在一定的阈值以内，曲线基本对齐，即认定该段时间序列具有对称性。例如，港口大型运输车往复作业，起重机抬臂和降臂工作等。

1.3 时间序列预测

时间序列预测任务是指从众多数据中挖掘出其蕴含的核心规律并且依据已知的因素对未来的数据做出准确的估计。

对于单步时间序列预测，使用历史观测值和协变量作为输入变量，来预测下一个时间步的观测值。这种预测过程被称为单步预测。

然而，在很多预测应用中，需要预测在未来一段时间内观测序列变化的趋势。例如，某一区域在未来一段时间的用电量、元器件在一段时间的退化趋势、股票价格在未来时间段中的变化趋势等。这种情况需要对未来多个时间步的时间序列进行预测。通常使用历史观测值和协变量作为输入，来预测未来 n 个时间步的预测值。这种时间序列预测被称为多步预测。

时间序列多步预测有五种策略，分别为直接多步预测(Direct Multi-step Forecast)、递归多步预测(Recursive Multi-step Forecast)、直接递归混合预测(Direct-Recursive Hybrid Forecast)、多输出预测(Multiple Output Forecast)、seq2seq 预测(seq2seq Forecast)。

第二章 时间序列预测方法

2.1 传统时间序列建模方法

传统时间序列预测方法主要是在确定时间序列参数模型的基础上，求解出模型参数，并利用求解出的模型完成未来预测工作。以下是一些典型的方法：

ARIMA 模型： ARIMA（Auto-Regressive Integrated Moving Averages）是一种被广泛使用的时间序列预测统计方法[4]。在处理过程中，首先需要对观测值序列进行平稳性检测，接着对其进行白噪声检验，计算 ACF（自相关系数）和 PACF（偏自相关系数）进行检验，进而进行预测。基于 ARIMA 方法还有许多改进算法，如 SARIMA 和 SARIMAX 等[5]。

Holt-Winters 方法： Holt-Winters 法扩展了传统的 Holt 方法来捕捉季节性[6]。Holt-Winters 方法包括预测方程和设置三个平滑方程中参数 α 、β 和 γ 分别用于水平、趋势和季节性分量预测。

对于处理简单的时序预测问题，传统时序模型可以发挥较大的优势。然而，如果变量和维度过多，变化模式过于复杂，传统时序模型就显得力不从心。在这种情况下，更复杂的模型可能更为适用。

2.2 基于时间序列分解的分析法

时间序列分解法一直是时间序列分析中非常有用的方法。该方法认为一个时间序列往往是以下多种变化形式的叠加或耦合：

长期趋势(Secular Trend, T)：在较长时期内持续发展变化的整体趋势或状态。

季节变动(Seasonal Variation, S)：由于季节变化引发的时间序列规则变化。

循环波动(Cyclical Variation, C)：以若干年（或周期）为期限，不具严格规则的周期性连续变动。

不规则波动(Irregular Variation, I)：由于各种偶然因素对时间序列发展造成的影响。

根据不同的应用场景，时间序列可以按照加法原则和乘法原则进行分解。加法模型中成分之间是相互独立的，某种成分的变动并不影响其他成分的变动，各个成分都用绝对量表示，并且具有相同的量纲。乘法模型中四种成分之间保持着相互依存的关系，一般而言，长期趋势用绝对量表示，具有和时间序列本身相同的量纲，

其他成分则用相对量表示。

一种典型的时间序列分解方法是 Facebook 开源的 Prophet 模型[7]，它将时间序列分解为趋势项、季节性项（周或月）、节假日项以及噪音项相加。

时间序列分解的方法更重要的是提供了一种对于时间序列的分析思路。在后续的机器学习和深度学习方法中，很多都按照这种分解思路对于时间序列预测问题进行设计和规划。

2.3 基于机器学习的方法

时间序列数据预测工作本质上与机器学习方法中的回归分析之间存在着紧密的联系。机器学习算法按照实现方法主要分为以下几类：

支持向量回归(SVR)：基于统计学习理论，具有很好的泛化能力。SVR 通过核函数方法将输入映射到高维空间，不增加计算复杂性，并有效克服维数灾难问题。在时间序列预测中，使用 SVR 称为支持向量回归(SVR)，对非线性时间序列具有稳定的预测能力[8]。SVR 通过非线性函数变换将输入映射到高维特征空间，然后在此高维特征空间中找到能够准确表明输出数据和输入数据之间关系的函数。

梯度渐进回归(GBRT)：是一种将梯度下降方法引入解决回归问题的方法。GBRT 算法通过利用损失函数的负梯度进行计算和迭代，最终得到最佳模型。该算法的目的是找到一个函数，使得该函数对应的损失函数期望最低[9]。GBRT 算法实际上是利用多个基函数的累加展开式对该函数进行逼近的方法。它对异常值的鲁棒性较高，广泛用于时间序列预测。

隐马尔可夫模型(HMM)：是一种统计学模型，为多变量时间序列预测建模提供了概率框架。HMM 是双随机过程，包括具有一定状态数的隐马尔可夫链和显示随机函数集。隐马尔可夫链不可观测，但可以通过产生一系列观测结果的一组过程来估计[10]。HMM 是一种最简单的动态贝叶斯网络，其他模型在动态贝叶斯网络框架内泛化了基本的 HMM，允许更多隐藏状态变量。 HMM 提供了一个概率框架，计算观察序列的发生概率，并选择最能解释观测结果的状态序列。

2.4 基于深度学习的方法

随着深度学习在计算机视觉和自然语言处理方面取得显著成就，深度学习方法也逐渐引入到时间序列预测应用中。通过构建各种网络结构，深度神经网络能够更好地进行高维数据表征，从而减少对手动特征工程和模型设计的需求。通过定义损失函数，深度神经网络可以更方便地进行端到端的训练。

深度神经网络通过多个非线性层来构建以往时间序列特征表示，从而学习时间序列内部变化规律。在神经网络的学习过程中，可以认为将相关历史时序信息编码到隐变量中，最终成为了隐变量和当前输入的输出。

针对不同的应用场景，在深度学习中隐藏层采用了不同的实现形式，从而采用了不同的神经网络架构来进行实现。

2.4.1 基于卷积神经网络的方法

卷积神经网络（CNN）在计算机视觉方面得到广泛应用，能够提取跨空间维度不变的局部特征。同样，它也可以在时间维度提取出局部特征，通过多层隐层逐渐在上一层进行汇总，从而获得过去时间段较长序列的隐藏信息。

改进前的传统卷积神经网络中，通过多层隐层汇聚，输入信息和最终输出预测结果的关联有限，输出信息难以关联到之前更远的时间序列信息，导致传统卷积神经网络在时间序列预测能力上表现有限。

为了扩大隐藏层的感受范围，DeepMind 提出了基于空洞因果卷积的语音生成模型 WaveNet[11]。后续专家们发现这种方法也适用于时间序列的预测。膨胀卷积可以解释为对于本层特征进行下采样的卷积，即降低分辨率以合并来自更久远信息。通过增加每一层的膨胀率，膨胀卷积可以逐渐聚合不同时间块的信息，有效地利用到时间序列中长远的历史信息[12]。

时间卷积网络（TCN）将序列看作一个一维对象帧，通过迭代多层卷积来捕捉长期关系[13-14]。TCN 利用因果卷积、空洞卷积以及残差卷积的跳层连接，适应时序数据的时序性并提供更宽广的时域感受野用于时序建模。 TCN 网络结构允许并行计算输出，提高了运行效率。基于这个思想，出现了一系列基于 TCN 网络的改进网络。

SCINet 采用层次化卷积网络结构，以不同的时间分辨率迭代地提取和汇聚特征，学习具有增强可预测性的有效表示[15]。整个 SCINet 结构是一个二叉树，在每个 SCI-Block 中，时间序列被分为两个部分，随着二叉树深度的增加，更精细的时序信息也会被提取出来。通过这种方式，时间序列中短期和长期的依赖关系都能被网络学习到。

2.4.2 基于循环神经网络的方法

循环神经网络（RNN）一直被用于序列建模，在各种自然语言处理任务上取得了很好的效果。RNN 网络的核心是可以学习到预测前所有时间序列内部隐藏状态，作为过去信息的特征表征，并结合当前的输入给出下一步预测的结果。在每个时间步，可以使用新的观测值不断地递归更新隐层状态。因此，在深度学习方法中，基

于 RNN 的网络架构最早被应用于时间预测场景。

早期的 RNN 变体在学习数据中的长期串行依赖关系时可能会受到梯度爆炸和消失问题的限制。为了解决这个问题，引入了长短期记忆（LSTM）网络[16]。LSTM 网络的设计灵感来自于计算机的逻辑控制门，通过各种门控制记忆元素，例如使用输出门控制从单元中输出序列，使用输入门决定何时将数据读入单元，使用遗忘门管理重置单元的内容。LSTM 通过训练得到的权重可以决定何时记忆或忽略隐状态中的输入，因此成为基于 RNN 的时间序列预测的重要组成部分。

DeepAR 网络采用 LSTM 模型解决时间序列预测问题[17]。在训练阶段，每个时刻传入上一时刻的真实值和协变量输入，经过 LSTM 单元后预测下一个时刻的值。在预测阶段，将模型在上一个时刻的预测值作为输入，替代训练过程中上一个时刻真实值的输入。

基于循环网络的 Deep State Space 模型包含一种基于状态空间转换的预测方法[18]。该模型预估了连续两个隐状态的关系，实现了从当前的隐状态到当前时刻预估结果的预测。相比之前的方法，该模型不需要在预测过程中输入上一个时刻的真实值或预测值，通过隐状态建立连续两个时刻的联系，解决了训练和预测不一致的问题。

为了克服训练模型时可能产生的大量复制数据和占用大量内存资源的问题，交叉式训练方法被提出[19]，该方法在不复制数据的情况下达到了滑动窗相似的训练效果，很好地兼顾了短期和长期预测。

此外，一种能结合指数平滑和循环神经网络的混合模型被提出[20]，兼容了各自的建模优势，有助于提升时间序列预测性能，取得了良好效果。

2.4.3 基于自注意力网络的方法

自注意力网络最早产生于自然语言处理任务，其中 Transformer 是 2017 年的论文《Attention is All You Need》中提出的自注意力模型[21]。这篇论文通过实验针对机器翻译场景，取得了当时最好的效果。由于 Transformer 的编码器部分采用了并行计算方式，训练和推理时间大大缩短，为序列化信息处理开辟了新的思路。目前在自然语言处理各业务主流的语言模型，如 GPT、BERT 等，都是基于 Transformer 模型。由于与自然语言处理任务的相似性，Transformer 的方法也很快被应用到时间序列预测任务上。

在循环神经网络（RNN）中，分析时间序列需要按照依次迭代的顺序处理从 t-n 到 t 的所有信息。当两个时间点相距较远时，它们之间的隐层联系通常较差，而且由于其顺序性处理效率也较低。基于注意力机制的模型可以实现输入时间序列中每个

单元之间的两两关联，对于下层特征通过关联权重加权的表征，提供给上一层进行处理。自注意力机制可以更好地实现时序的上下文信息交互。

文献 [22] 采用了与 GPT 相似的 Transformer 结构尝试了时间序列预测任务，并取得了较好的效果。Transformer 模型具有提高预测能力的潜力，但同时也存在计算量大、内存占用高和编码器-解码器架构等限制，使其无法直接应用于更长时间序列预测问题。文献 [23] 提出了卷积自注意力，通过在自注意力层中使用因果卷积生成查询（Queries）和键（Key），降低了计算复杂度。Informer[24]通过选择 O(log L) 占主导地位的查询，从而在计算复杂度上实现了改进，并通过生成式解码器直接产生长期预测序列，避免了使用单步前向预测时的累积误差。

另一方面，学者们在时间序列建模中探索频域自注意力机制。Autoformer[25]设计了一种短期趋势分解架构，其自相关机制用作注意力模块。不同于之前的注意力机制，它测量输入信号之间的时延相似性，并聚合前 k 个相似的子序列以产生具有 O(L log L)复杂度的输出。FEDformer[26]通过傅里叶变换和小波变换在频域中应用注意力操作，通过随机选择固定大小的频率子集来实现线性复杂度。

2.4.4 基于图神经网络的方法

多元时间序列预测中，变量之间存在相互依赖，而现有方法无法完全利用变量之间的潜在空间依赖性。近年来，图神经网络（Graph Neural Network, GNN）在处理关系依赖方面表现出了很强的能力。通常，通过图结构学习层定义信息传播的图结构，然后进行卷积操作，通过捕获时间序列内的空间和时间依赖性，对多变量时间序列进行预测。

文献 [27] 提出了一个为多元时间序列数据设计的通用图神经网络框架。通过图学习模块自动提取变量之间的关系，并将变量信息整合到其中。在网络中，图卷积模块与时间卷积模块相互交错，分别捕获空间和时间相关性。使用混合跳跃传播层来捕获时间序列内的空间和时间依赖性。

文献 [28] 在多变量时间序列预测中使用图神经网络，采用了 N 个节点的全连接图，按照内部相关性映射为 K 个节点连接图，从而减少了图神经网络的节点规模，将图模型的复杂度由 O(N^2)转化为 O(NK)，实现了精度和计算复杂度之间的平衡。

文献 [29] 提出了一种新颖的图神经网络架构 Graph WaveNet，用于时空图建模。通过引入自适应的图建模方法并通过节点嵌入来学习，可以精确地捕捉数据中隐藏的空间依赖性。同时，通过空洞因果一维卷积模型替代 RNN 模型，更容易提取长期时序关系。

2.4.5 基于残差的全连接网络方法

文献 [30] 提出了 Nbeats 模型，该模型内部结构中没有 RNN、CNN 或 Attention，网络全部由全连接层组成，在时间序列预测方面取得了很好的效果。

Nbeats 的核心思路是通过多个全连接卷积块（FC Blocks）串联，每个 FC Block 学习输入时间序列的一部分信息。进入下一个 FC Block 的输入会去掉之前 Block 已经学到的信息，后续的 Block 只关注学习于之前 Block 未学到的信息，类似于 GBDT 不断拟合的思路。在预测过程中，将综合各个 Block 的预估结果得到最终预估结果。通过这种方式，模型每层更加关注前层无法正确拟合的残差，也将时间序列进行了从趋势到细节的逐层分解。

具体而言，Nbeats 的模型结构包括多个 Stack，每个 Stack 包括多个 FC Block，每个 FC Block 为最基础的结构模块，由多个全连接层组成。

为了考虑到时间序列预测问题中的外部特征，如日期信息、节日信息、属性信息等，NBEATSx[31] 提出了可以引入外部特征的 Nbeatsx 网络。与初版 Nbeats 的主要区别是引入了外部特征 x。GAGA[32]将 Nbeats 扩展到了时空预测领域，能够处理存在空间关系的多个时间序列的建模。在先前的模型增加了 Time Gate 和 Graph Gate 等网络结构，嵌套到先前的网络中。

相比其他时间序列预测模型，Nbeats 独创了一种全部为全连接的主体网络架构。通过对时间序列进行由粗到细的序列信息分解，不断进行细化，实现准确的时间序列预测。

第三章 总结与展望

本文首先详细介绍了时间序列预测的相关背景知识，其次对于早期用于时间序列预测的诸多方法进行了介绍，其中包括传统方法和机器学习的方法，同时也在此基础上分析出了这些方法的继承点和不足的地方，最后重点介绍和对比各基于深度学习的不同模型设计的方法。

虽然基于深度学习的时间序列预测方法取得了显著的发展，但未来的预测应用场景对其提出了更高的要求。为进一步研究这一技术，几个主要的研究方向如下：

1、整合领域知识：针对不同的使用领域，深度学习模型需要充分整合各自领域的知识，以获得更高的精度和效率。这意味着在建立模型时要考虑特定领域的特殊性，充分利用领域专业知识。

2、因果性推断：目前时间序列预测主要通过序列内部的相关性分析实现，未来的应用中需要更多关注因果性推断。模型的预测可解释性需要进一步加强，确保用户能够理解模型的决策过程。

3、异常点处理：在实际预测的过程中，获得的数据通常存在异常点。为了提高模型的鲁棒性，需要在预测过程中忽略这些异常点，以减小预测方法的误差。

4、在线学习：目前深度学习方法主要通过线下收集数据进行批处理训练，未来需要发展增量学习的方法，使模型能够对时间序列数据进行在线分析预测。这有助于模型实时地适应新数据和变化。

5、采用随机自然启发优化算法优化深度学习模型的超参数： 随着深度学习算法的复杂性增加，需要处理的超参数也越来越多，而超参数的选择直接影响算法的性能。随机自然启发优化算法灵感来自群体智能和自然现象，通过随机生成可解空间内的一定数量解，并在算法的各个阶段重复寻找全局最优解，可用于提升深度学习模型的预测能力，成为未来研究热点之一。

6、研究适合时间间隔不规则的小数据集的网络架构： 现有的 Transformer 模型在周期性好的大数据集上表现出色，但在小数据量和时间间隔不规则的数据集中表现不理想。为了解决过拟合问题，未来的研究可以在模型架构中引入重采样、插值、滤波或其他方法，为处理时间序列数据和任务特征提供新思路。

7、引入图神经网络（GNN）用于多变量时序预测建模： 由于多变量时序预测任务的潜在变量相关性复杂，而现实世界中的数据相关性是变化的，因此 GNN 可以更好地处理多变量时序预测。最近，采用时间多项式图神经网络的方法在短期和长期多变量时序预测上取得了先进的水平，因此 GNN 对于多变量时序预测的建模能力值得深入研究。

8、研究同时支持精确形状和时间动态的可微损失函数作为评价指标： 在时间序列预测领域，已经使用了许多测量度量，但现有的损失函数往往对形状和时间延后失真不具有不变性。未来的研究方向可以探索同时支持精确形状和时间动态的可微损失函数，以更全面地评价模型的性能，帮助生成更及时、更稳健和更准确的预测。

通过对这些方向的深入研究，基于深度学习的时间序列预测方法将能够更加安全可靠、高效灵活地解决实际问题，实现更好的实际应用效果。

结束语

随着信息时代的进一步发展，在实际生产生活中，任何领域都必将产生大量的时序数据，从这些时间序列数据中发现变化规律，并基于这些变化规律对未来的生产生活进行预测的任务背后蕴含着巨大的经济价值和科研价值。研究该时间序列数据预测任务的方法从传统统计学到机器学习，再到目前深度学习的方法，预测的精准度和速度在一步步提高。本文详细回顾了这一研究方法变化过程，梳理出背后的发展规律，在此基础上对于后续的研究进行了展望。

参考文献

[1] 次必聪，张品一. 基于 ARIMA-LSTM 模型的金融时间序列预测[J]. 统计与决策，2022，38（11）：145-149. DOI: 10.13546/j.cnki.tjyjc.2022.11.029.

[2] 车畅畅，王华伟，倪晓梅，等. 基于 1D-CNN 和 Bi-LSTM 的航空发动机剩余寿命预测[J]. 机械工程学报，2021，57（14）：304-312. DOI: 10.3901/JME.2021.14.304.

[3] 罗晓萌，李建斌，胡鹏. 基于时间序列预测的电子商务库存优化策略[J]. 系统工程，2014，32（6）：91-98.

[4] 张 G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing，2003，50：159-175. DOI: 10.1016/S0925-2312(01)00702-0.

[5] 李少雄，李本光. 基于 SARIMA 模型和 X-12-ARIMA 季节调整方法预测的比较[J]. 统计与决策，2018，34（18）：39-42. DOI: 10.13546/j.cnki.tjyjc.2018.18.008.

[6] Chatfield C. The holt-winters forecasting procedure[J]. Journal of the Royal Statistical Society，1978，27（3）：264-279. DOI: 10.2307/2347162.

[7] Taylor S J, Letham B. Forecasting at Scale [J]. PeerJ Preprints，2017，5：e3190v2. DOI: 10.7287/peerj.preprints.3190v2.

[8] 陈荣，梁昌勇，谢福伟. 基于 SVR 的非线性时间序列预测方法应用综述[J]. 合肥工业大学学报 (自然科学版)，2013，36（3）：369-374. DOI: 10.3969/j.issn.1003-5060.2013.03.025.

[9] Elsayed S, Thyssens D, Rashed A, et al. Do we really need deep learning models for time series forecasting?[Z]. arXiv: 2101.02118, 2021. DOI: 10.48550/arXiv.2101.02118.

[10] Zahari A, Jaafar J. A novel approach of hidden Markov model for time series forecasting[C]//Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. Bali, Indonesia: ACM, 2015: 91. DOI: 10.1145/2701126.2701179.

[11] Van den Oord A, Dieleman S, Zen H G, et al. WaveNet: A generative model for raw audio[C]//The 9th ISCA Speech Synthesis Workshop. Sunnyvale, CA, USA: ISCA, 2016: 125.

[12] Borovykh A, Bohte S, Oosterlee C W. Conditional time series forecasting with convolutional neural networks[Z]. arXiv: 1703.04691, 2017. DOI: 10.48550/arXiv.1703.04691.
[13] Borovykh A, Bohte S, Oosterlee C W. Dilated convolutional neural networks for time series forecasting[J]. Journal of Computational Finance，2019，22（4）：73-101. DOI: 10.21314/JCF.2019.358.
[14] Chen Y T, Kang Y F, Chen Y X, et al. Probabilistic forecasting with temporal convolutional neural network[J]. Neurocomputing，2020，399：491-501. DOI: 10.1016/j.neucom.2020.03.011.
[15] Liu M H, Zeng A L, Chen M X, et al. SCINet: Time series modeling and Forecasting with sample convolution and interaction[Z]. arXiv: 2106.09305, 2021. DOI: 10.48550/arXiv.2106.09305.
[16] Hossain M S, Mahmood H. Short-term load forecasting using an LSTM neural network[C]//2020 IEEE Power and Energy Conference at Illinois (PECI). Champaign: IEEE, 2020. DOI: 10.1109/PECI48348.2020.9064654.
[17] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting，2020，36（3）：1181-1191. DOI: 10.1016/j.ijforecast.2019.07.001.
[18] Rangapuram S S, Seeger M W, Gasthaus J, et al. Deep state space models for time series forecasting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018: 7796-7805.
[19] Wen R F, Torkkola K, Narayanaswamy B, et al. A multi-horizon quantile recurrent forecaster[Z]. arXiv: 1711.11053, 2017. DOI: 10.48550/arXiv.1711.11053.
[20] Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting[J]. International Journal of Forecasting，2020，36（1）：75-85. DOI: 10.1016/j.ijforecast.2019.03.017.
[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 6000-6010.
[22] Wu N, Green B, Ben X, et al. Deep transformer models for time series forecasting: the influenza prevalence case[Z]. arXiv: 2001.08317, 2020. DOI: 10.48550/arXiv.2001.08317.
[23] Li S Y, Jin X Y, Xuan Y, et al. Enhancing the locality and breaking the memory bottle

neck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran Associates Inc., 2019: 471.
[24] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Event: AAAI Press, 2021: 11106-11115.
[25] Wu H X, Xu J H, Wang J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event, Curran Associates, Inc., 2021, 34: 22419-22430.
[26] Zhou T, Ma Z Q, Wen Q S, et al. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International Conference on Machine Learning. Baltimore: PMLR, 2022: 27268-27286.
[27] Wu Z H, Pan S R, Long G D, et al. Connecting the Dots: multivariate time series forecasting with graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event, CA, USA: ACM, 2020: 753-763.
[28] Satorras V G, Rangapuram S S, Januschowski T. Multivariate time series forecasting with latent graph inference[Z]. arXiv: 2203.03423, 2022. DOI: 10.48550/arXiv.2203.03423.
[29] Wu Z H, Pan S R, Long G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China: IJCAI.org, 2019: 1907-1913.
[30] Oreshkin B N, Carpov D, Chapados N, et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting[C]//8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net, 2020.
[31] Olivares K G, Challu C, Marcjasz G, et al. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx[Z]. arXiv: 2104.05522, 2021. DOI: 10.48550/arXiv.2104.05522.
[32] Oreshkin B N, Amini A, Coyle L, et al. FCGAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence. Virtual Event, AAAI Press, 2021: 9233-9241.